DOCUMENT RESUME

ED 344 836 SO 030 252

AUTHOR Fisher, Michelle B.
TITLE A Pilot Study on the Extent to Which a Standardized Subject-Matter Test Assesses Critical Thinking Abilities. Critical Thinking Reports Number 5.
INSTITUTION Illinois Univ., Urbana. Dept. of Educational Policy Studies.
PUB DATE 89
NOTE 21p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Critical Thinking; *Educational Research; High Schools; Social Studies; *Standardized Tests; Testing; Test Interpretation; *Test Reviews; Thinking Skills
IDENTIFIERS *Sequential Tests of Educational Progress

ABSTRACT

This document is a pilot study designed to examine the extent to which currently-used subject-matter standardized tests already test for critical thinking in order to determine if there is a need for the construction of new tests. The study also is intended to be a step toward devising a reliable procedure and scoring key that will enable future analyses of both subject-specific and general knowledge tests. The social studies subtest of the Sequential Tests of Educational Progress was examined and scored based on the number of critical thinking abilities that were tested for by each item. It was found that two-thirds of the test questions did not test for a single critical thinking ability, and that the remaining portion of the test only reached five of the twelve identified critical thinking abilities. A list of references and two appendices are included. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED344836

A PILOT STUDY ON THE EXTENT TO WHICH A STANDARDIZED SUBJECT-MATTER TEST ASSESSES CRITICAL THINKING ABILITIES

BY

MICHELLE B. FISHER

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MICHELLE B. FISHER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CRITICAL THINKING REPORTS

NUMBER 5

General Editor of Critical Thinking Reports
Robert H. Ennis

Published by

The Illinois Critical Thinking Project

Department of Educational Policy Studies
University of Illinois at Urbana-Champaign
Urbana, Illinois
1989

SO 020 252

BEST COPY AVAILABLE

A PILOT STUDY ON THE EXTENT TO WHICH A STANDARDIZED SUBJECT-MATTER TEST ASSESSES CRITICAL THINKING ABILITIES

BY

MICHELLE B. FISHER

CRITICAL THINKING REPORTS

NUMBER 5

General Editor of Critical Thinking Reports
Robert H. Ennis

Published by

The Illinois Critical Thinking Project

Department of Educational Policy Studies
University of Illinois at Urbana-Champaign
Urbana, Illinois
1989

Author Notes

I wish to thank James Kotowski and Robert H. Ennis for their careful reading of this paper and their valuable suggestions.

A limited supply of copies of this report is available at $5.00 per copy plus $2.00 postage and handling. Requests should be sent to the Illinois Critical Thinking Project, 1310 South Sixth Street, Champaign, Illinois 61820.

A PILOT STUDY ON THE EXTENT TO WHICH A STANDARDIZED SUBJECT-MATTER TEST ASSESSES CRITICAL THINKING ABILITIES

Michelle B. Fisher
Illinois Critical Thinking Project
Department of Educational Policy Studies
University of Illinois at Urbana-Champaign

Recently, The Nation's report card (1987) called for the assessment of "higher order thinking skills." In the review section of this report, Robert Glaser, speaking for the National Academy of Education, states the following (p. 54):

> To remove the assessment of these [higher order thinking] skills from the context of their subject matter seems likely to trivialize them and over time may lead to their neglect in the instruction of the core subjects. Indeed, we find it hard to imagine good teaching and assessment without an emphasis on the ability to reason effectively within each domain.

Included as higher order thinking skills are "gathering relevant information, evaluating evidence, weighing alternative courses of action, and articulating reasoned arguments" (p. 54). Such skills as these are consistent with and expanded upon in Ennis' critical thinking taxonomy (1987, pp. 12-15):

> Dispositions: 1) seek a clear statement of the thesis or question; 2) seek reasons; 3) try to be well-informed; 4) use credible sources and mention them; 5) take into account the total situation; 6) try to remain relevant to the main point; 7) keep in mind the original and/or basic concern; 8) look for alternatives; 9) be open-minded; 10) take a position (and change a position) when the evidence and reasons are sufficient to do so; 11) seek as much precision as the subject permits; 12) deal in an orderly manner with the parts of a complex whole; 13) use one's critical thinking abilities; and 14) be sensitive to

the feelings, level of knowledge, and degree of sophistication of others.

Abilities: 1) identifying and focusing on a question; 2) analyzing arguments; 3) asking and answering questions of clarification and/or challenge; 4) judging the credibility of a source; 5) observing and judging observation reports; 6) deducing and judging deductions; 7) inducing and judging inductions; 8) making and judging value judgments; 9) defining terms and judging definitions; 10) identifying assumptions; 11) deciding on an action; and 12) interacting with others.

Since Ennis' taxonomy is more systematic and complete than the above list of "higher order thinking skills," we will use it for the purposes of this study, with a focus on the abilities.

We endorse the National Academy of Education's concern for the assessment of critical thinking within subject-matter areas to ensure its being taught in the classroom. Accordingly, we are interested in finding out the extent to which currently-used subject-matter tests already test for critical thinking in order to determine if there is a need for the construction of new tests that better accomplish this task.

What we have set out to do in this study is to explore the problems and prospects for examining the degree of critical thinking tested for by subject-matter tests. The procedure we used was exploratory and is subject to further revision--as is the critical thinking taxonomy that we used. This pilot study is a step toward devising a reliable procedure and scoring key that will enable future analyses of both subject-specific and general-knowledge tests (Fisher, Kotowski, & Ennis, in process). With such a procedure, we can fulfill our immediate goal of

ascertaining the extent to which current subject-matter tests assess critical thinking. With an idea of what exists now, we will better know what needs to be done in the future in the area of assessment of thinking skills.

Procedure

We selected the Sequential Tests of Educational Progress (1957) from a list of achievement tests used in grade 10, cited in Performance profiles: Illinois schools report to the public (Illinois State Board of Education, 1987, Table 5, p. 31). The fact that this test is published by the largest testing organization in America, Educational Testing Service, aided in our selection. We selected only one test so that we could keep the exploratory task as simple as possible. We then narrowed our selection to the social studies subtest (STEP-SS) since it was a subject with which all three scorers felt fairly comfortable. The test was then examined and scored in the following manner: Scorer #1 administered the test (both parts I and II) to herself, scoring each question according to those critical thinking abilities it tested for, using Ennis' critical thinking taxonomy (1987). Scorers #2 and #3 repeated this procedure. Then, all three scorers went through the test together and compared their prior judgments of each test item. Where test items were scored differently by each scorer, a discussion ensued with a final judgment being mutually agreed to as to which critical thinking abilities were tested for by each test item. We have not yet

determined an inter-rater reliability measure. This is beyond the scope of this study, which is intended to describe the results of an exploratory procedure for scoring tests for critical thinking. An inter-rater reliability measure will be included in later analyses which make use of the standardized scoring key (see Fisher, Kotowski, & Ennis, in process).

Results

The test relied most heavily on knowledge of subject matter (viz., social studies) and literal reading comprehension (i.e., understanding the meaning of the words and sentences used throughout the test). We also noticed that many of the items could be answered using test-wiseness (without recourse to subject matter knowledge). Although test-wiseness does involve critical thinking, it does not involve critical thinking <u>in the subject area</u>; hence, we did not include it in our scoring.

A brief summary of our results are as follows: 47 of the 70 questions (67%) did not test for any single critical thinking ability. Of the remaining 23 items (33%), 14 tested for only 1 ability, 8 tested for a total of 2 abilities, and 1 tested for a total of 3 abilities (see Figure 1).

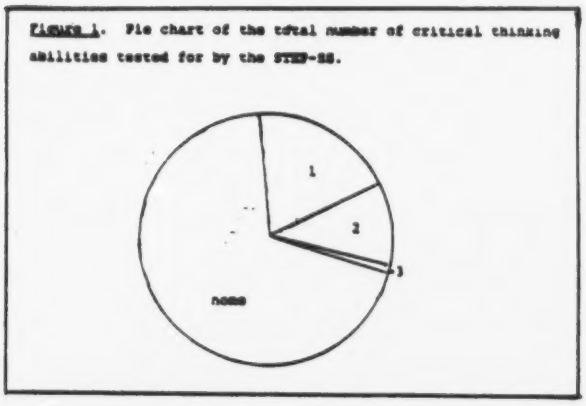


Figure 1. Pie chart of the total number of critical thinking abilities tested for by the STEP-SS.

Test items were scored each time they tested for a critical thinking ability. Since some items tested for more than one ability, 9 questions were counted more than once. Seven out of the twelve critical thinking abilities were not assessed at all. Of the five abilities that were tested for, these are the percentage of items that did so: 1.4% clarification, 1.4% value judgments, 8.6% deduction, 11.4% induction, and 24.3% observation (see Table 1 and Figure 2 next page). Our results for each test item of the STEP-SS are provided in Appendixes A and B.

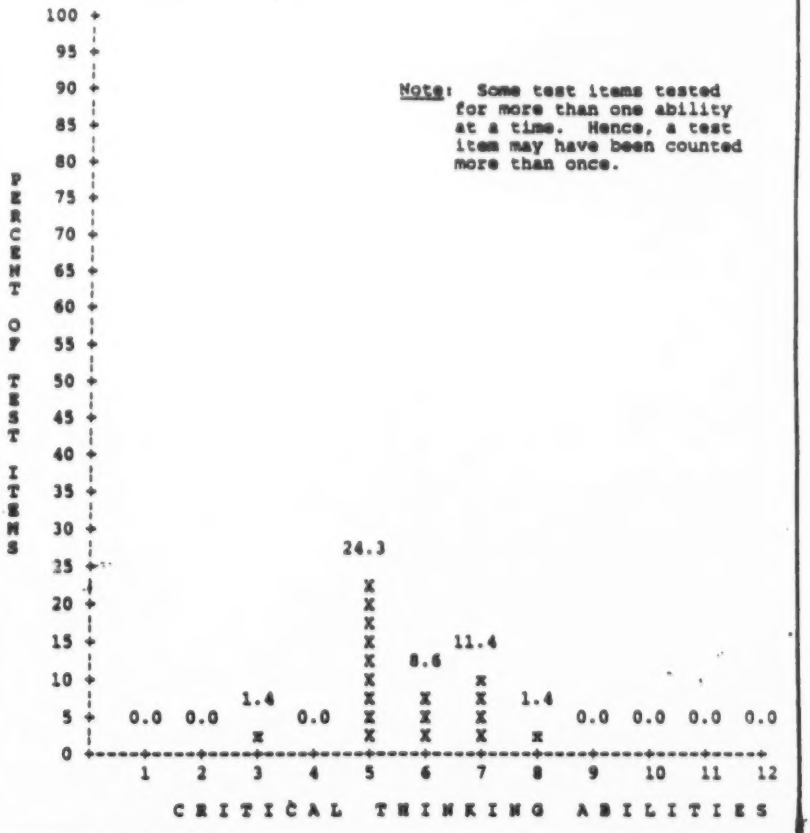
Discussion

It must first be noted that, due to the fact that we analyzed only one subtest of one subject-matter test, any conclusions or generalizations made on the basis of this study should be at most very tentative. We should also point out that test takers may answer test items using different approaches or combinations of approaches, e.g., relying on memorized factual knowledge, critical thinking, and/or test-wiseness. We cannot know which approach(es) each or most test takers use for each item. If anything, our scoring was generous in assigning critical thinking to test items because it is very probable that test takers would use the easiest method they know for answering the question--viz., memorized factual knowledge; if such knowledge is not readily available, the test taker would either use test-wiseness strategies (which involve looking at the form rather than the content of the test item in order to answer it)

Table 1. Extent to which each critical thinking ability was assessed by the STEP-SS.

Abilities	Test items that tested for each ability	
	Frequency	Percentage
1) identifying and focussing on a question	0/70	0.0
2) analyzing arguments	0/70	0.0
3) asking and answering questions of clarification and/or challenge	1/70	1.4
4) judging the credibility of a source	0/70	0.0
5) observing and judging observation reports	17/70	24.3
6) deducing and judging deductions	6/70	8.6
7) inducing and judging inductions	8/70	11.4
8) making and judging value judgments	1/70	1.4
9) defining terms and judging definitions	0/70	0.0
10) identifying (used and needed) assumptions	0/70	0.0
11) deciding on an action	0/70	0.0
12) interacting with others	0/70	0.0

Figure 2. Bar graph of the percentage of test items that tested for each critical thinking ability (as listed in Table 1).



or a combination of critical thinking and subject matter knowledge. It is important to remember that critical thinking is not divorced from content and therefore cannot be tested for purely; hence, items that are scored as testing for critical thinking also test for subject-matter knowledge. We realize there are problems such as those mentioned above which make our scoring results tentative ones. Given these constraints, however, we can examine the results we have attained thus far for some implications.

Although Ennis' taxonomy does include a list of critical thinking dispositions, we regrettably did not score the STEP-SS for the extent to which these were tested for. It is nevertheless apparent that some dispositions are necessary to do well on the test, such as trying to be well-informed and dealing in an orderly manner with the parts of a complex whole; on the other hand, this test did not assess dispositions such as looking for alternatives and being open-minded (as discussed below). Evaluation of such dispositions seems to require more direct interaction between those evaluating and those evaluated than found in the test-taking situation--the classroom seems to be a better forum for such assessment.

Some test items (9) assessed more than one critical thinking ability at a time. When a greater number of critical thinking abilities must be integrated to answer a test item, more sophisticated thinking is often required. Hence, these items can generally be considered to be more sophisticated ones in testing

for critical thinking.

The most extensively assessed ability was observation. A total of 24.3% of the items tested for observation, most of which were at a moderate level of sophistication where further interpretation and/or use of other critical thinking abilities was required. These test items included those which asked the test taker to interpret maps, graphs, and political cartoons. Without such devices, this particular ability would not be assessed.

The next two most extensively assessed abilities were induction and deduction, tested for by 11.4% and 8.6% of the test items, respectively. This finding is a positive one, in comparison to the extent to which other abilities were assessed by this test. It should be noted here that we make no claims about the optimum extent to which tests should assess each ability; this is an open question.

Asking and answering questions of clarification and making or judging value judgments were tested by only one item each. Moreover, many critical thinking abilities were not tested at all, including: identifying and focussing on a question, analyzing arguments, judging the credibility of a source, defining terms and judging definitions, identifying assumptions, deciding on an action, and interacting with others. It is very likely that, had this test used essay questions, some of these abilities might have been included.

A number of items suggested critical thinking, but at such a

low level of sophistication that we hesitated to include these items in our scoring. In particular, we found a number of low-level deduction and low-level induction items where all that was necessary was "exemplification:" identifying an example of a general principle or concept (and vice versa). Exemplification embodies the lay conceptions of deduction and induction (two of the abilities listed in Ennis' taxonomy), where deduction involves going from a general principle to a specific case and induction involves going from a number of specific cases to a general principle. It is debatable, however, whether or not exemplification should be included as part of the critical thinking taxonomy. The problem with including exemplification is that it is often so basic as to not require much thinking, let alone <u>critical</u> thinking. For example, sample item 0 in the directions for part one of the STEP-SS (1957, p. 1) is an illustration of how basic this ability can be:

> 0 Which one of the following is an animal?
> A Bed
> B Dog
> C Chair
> D Box

This is a clear instance of exemplification that does not call for critical thinking. More complex items that test for exemplification include the following (p. 3):

> The following statements describe life during various periods in the development of Western civilization.
>
> I: The revenues which the landowners collect from their tenants are directed toward no economic purpose. They are scattered in alms, in the building of monuments, in the purchase of works of

art. Wealth, if one may so term it, is fixed motionless in the hands of an aristocracy.

II: Unrestrained competition sets capitalists to struggling with each other and soon arouses resistance among the proletariat that they are exploiting. And at the same time that resistance arises to confront capital, the latter, itself suffering from the abuse of freedom which had enabled it to rise, compels itself to discipline its affairs.

III: The state strives to increase its means of production, to close its markets to its competitors, and to make the balance of trade incline in its favor. Some of the devices it uses are: protective duties and bounties, subsidies to manufacturers and national navigation, acquisition of transmarine colonies, creation of privileged commercial companies, and inspection of manufacturing processes.

IV: The animals they slaughtered and the fruits that they gathered were wild. Their homes were caves found ready for use. The first beginnings in the cultivation of grain and the domestication of animals were just appearing. This brought an increase in population and the rise of communal life and domestic arts.

9. Which statement suggests the organization of such companies as the Hudson Bay and East India companies?
A I B II C III D IV

10. Statement IV best describes life in
E Colonial America
F Medieval Europe
G Hellenistic Greece
H Neolithic Europe

11. In the last sentence of Statement II, to which of the following might the author be referring when he mentions "the abuse of freedom"?
A Antitrust legislation
B Labor unions
C Cutthroat competition
D High taxes

In these instances, the thinking necessary for answering the items seems somewhat more sophisticated, yet still lower order. In multiple-choice question tests, exemplification appears to

consist of subject matter knowledge and literal reading comprehension (and perhaps test-wiseness). Exemplification is assuredly a lower-order thinking skill within situations in which there is only one clear-cut right answer. In cases where the answer is controversial or unsettled, however, exemplification enlists the use of other critical thinking abilities and thereby becomes a more sophisticated skill. For example, judging whether or not a woman who undergoes an abortion is a murderer, whether or not a politician is a liberal, or whether or not supporting a group of rebels is an instance of furthering democracy are very complex questions that involve a high level of critical thinking. These are cases where the answer is not black and white, unlike questions that ask whether or not people living in caves is characteristic of the Neolithic Age. In summary, although exemplification can be a sophisticated skill, in this test it is not, because of the settled nature of the questions.

In general, this test did not allow for test takers to question the truth of statements or answer choices. A more critical stance towards the test information was not encouraged and test takers did not have the opportunity to disagree. This seems to reinforce the more general tendency of students believing everything the teachers (or texts) say to be true.

Furthermore, test takers are not encouraged to come up with alternative justifications since there is only one answer keyed as correct. Moreover, test takers may be penalized if they do since they may select a non-keyed answer based on their

alternative justifications. For example, item 8 (p. 2) may be answered in a number of alternative ways:

8 What is the likelihood that the problem depicted in the cartoon will exist and be significant ten years from now?
- E It is likely to exist, and it is likely to be significant.
- F It is likely to exist, but it is not likely to be significant.
- G It is not likely to exist, but it is likely to be significant if it does exist.
- H It is not likely to exist, and it is not likely to be significant.

Let us assume that the problem depicted in the cartoon is that the constitution does not equitably represent the population--that it favors rural areas over cities despite the greater number of people in urban areas. Two justifications could be given to support selecting two different answers: 1) Answer E: Farmers are likely to continue to be a strong lobbying force in the future thus perpetuating the problem depicted above; this is likely to be significant because people in cities will not

receive equitable federal programs and funding to deal with pervasive urban problems such as poverty and homelessness. 2) Answer G: People living in cities are beginning to unite to lobby for more equitable funding and programs thus making the problem depicted above unlikely to continue; but if it did exist, it would be significant for the same reason as in (1).

Coming up with alternative justifications (than those of the test constructors) may be indicative of different background knowledge and/or more imaginative or sophisticated level/depth of critical thinking. If this is the case, this test may be testing only a low or unsophisticated level of critical thinking. More specifically, disallowing for alternative justifications in a social studies test reinforces the idea that history is not subject to more than one interpretation--that of the social studies text and teacher.

In summary, two-thirds of the questions did not test for a single critical thinking ability, and only 5 of the 12 abilities were necessary for answering the remaining third of the test. Of those 5 abilities that were tested for, 2 of them were assessed by less than 2% of the test items, and 1 by less than 9% of the items. Given these facts, it can be concluded that the STEP-SS test only minimally incorporated critical thinking.

Conclusion

The results we have obtained in this study, again, are tentative (due to the provisional nature of our category scheme).

Further studies need to be undertaken before results can be generalized. However, if this test is any indication of other subject-matter tests, the current state of affairs is not a bright one. This should be of concern to us if we agree with the National Academy of Education in that it is necessary to incorporate higher-order thinking skills within the context of subject-matter tests to avoid the neglect and trivialization of such thinking skills in the classroom.

References

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), Teaching Thinking Skills: Theory and Practice (pp. 9-26). New York: Freeman & Co.

Fisher, M. B., Kotowski, J., & Ennis, R. H. (in process). A system for determining the critical thinking content in tests (Critical Thinking Report No. 6). Champaign, IL: Illinois Critical Thinking Project, University of Illinois.

Illinois State Board of Education. (1987). Performance profiles: Illinois Schools Report to the public; a summary and analysis of information from the 1986 report cards. Springfield, IL: Illinois State Board of Education.

The Nation's report card: Improving the assessment of student achievement. (1987). Cambridge, MA: National Academy of Education.

Sequential Tests of Educational Progress: Social Studies, Form 1A. (1957). Princeton, NJ: Educational Testing Service.

Appendix A

Scores for each test item in the Sequential Tests of Educational Progress: Social Studies, Form 1A, part I.

question #	critical thinking ability	# in taxonomy
I-1	----------	--
I-2	----------	--
I-3	----------	--
I-4	----------	--
I-5	induction, observation	7, 5
I-6	induction, deduction, observation	7, 6, 5
I-7	deduction	6
I-8	deduction, value judgment	6, 8
I-9	----------	--
I-10	----------	--
I-11	----------	--
I-12	----------	--
I-13	----------	--
I-14	----------	--
I-15	----------	--
I-16	----------	--
I-17	----------	--
I-18	----------	--
I-19	observation, induction	5, 7
I-20	observation	5
I-21	----------	--
I-22	----------	--
I-23	----------	--
I-24	----------	--
I-25	induction	7
I-26	clarification	3
I-27	induction, deduction	7, 6
I-28	----------	--
I-29	----------	--
I-30	----------	--
I-31	----------	--
I-32	observation	5
I-33	observation	5
I-34	observation	5
I-35	observation	5

Appendix B

Scores for each test item in the Sequential Tests of Educational Progress: Social Studies, Form 1A, part II.

question #	critical thinking ability	# in taxonomy
II-1	observation, induction	5, 7
II-2	observation, induction	5, 7
II-3	observation, deduction	5, 6
II-4	----------	--
II-5	induction	7
II-6	----------	--
II-7	----------	--
II-8	----------	--
II-9	----------	--
II-10	----------	--
II-11	----------	--
II-12	----------	--
II-13	observation	5
II-14	observation	5
II-15	observation	5
II-16	observation	5
II-17	----------	--
II-18	----------	--
II-19	----------	--
II-20	----------	--
II-21	----------	--
II-22	----------	--
II-23	----------	--
II-24	----------	--
II-25	observation	5
II-26	deduction, observation	6, 5
II-27	----------	--
II-28	----------	--
II-29	----------	--
II-30	----------	--
II-31	----------	--
II-32	----------	--
II-33	----------	--
II-34	----------	--
II-35	----------	--